

BAB I

PENDAHULUAN

1.1. Latar Belakang

Sarkasme yaitu komunikasi yang bersifat mengolok-olok atau sebuah sindiran namun menggunakan bahasa yang keras dan kasar [1]. Sarkasme ini seringkali ditemukan dalam sebuah konten yang beredar diinternet khususnya sosial media. Dampak yang ditimbulkan oleh kata-kata sarkasme ini sangat besar bagi masyarakat Indonesia dalam berbagai faktor terutama konten yang menyangkut pendidikan, berita, hiburan dan lain-lain yang menjurus ke hal negatif. Memang sulit bagi kita untuk menyaring konten-konten yang berbau sarkasme ini melihat begitu banyak konten yang tersebar luas diberbagai media sosial seperti Youtube, Instagram, aplikasi X dan lain-lain. Kasus bertutur kata tidak baik sering bermunculan pada aplikasi X, pada aplikasi X, teks ataupun sebuah konten yang dibuat oleh netizen atau pengguna bebas mengekspresikan apapun secara perasaan maupun secara emosional dalam setiap peristiwa[2].

Banyaknya pengguna internet di Indonesia menurut survey we are social pada tahun 2022 yaitu sekitar 191 juta jiwa. Jumlah ini meningkat 12,35% dibandingkan dengan tahun 2021 dengan hanya 170 juta jiwa dan saat ini diperkirakan meningkat [3]. Peningkatan ini menciptakan keadaan baru antara positif dan negatif yang perlu perhatian khusus untuk menekan dampak negatif dari internet agar dapat diminalkan terutama terhadap isu netizen tidak sopan di Indonesia yang tentunya akan menjadi masalah berbahaya. Netizen Indonesia masih banyak yang melakukan pelanggaran etika walaupun sudah ada netiket dan menduduki peringkat yang sangat rendah dalam etika yaitu peringkat 29 dari 32 negara yang meliputi etika komunikasi [4]. Menurut data dari APJII pada tahun 2020 mengenai cyberbullying 31,8% memilih untuk membiarkan, 6,7% yang melaporkan kepada pihak berwajib serta masyarakat yang menjawab tidak tau yaitu 40% [5]. Jika tidak dicegah ini akan berdampak buruk bagi masyarakat banyak. Salah satu bentuk pelanggaran etika yaitu sarkasme, jika terus berkembang akan berdampak pada psikis seperti sakit hati, mental, dan rasa kecewa bagi korban maupun masyarakat yang lainnya.

Umumnya bahasa sarkasme digunakan untuk mengejek bahkan menjatuhkan mitra tutur. Netizen yang berkomentar yang bermaksud untuk mengkritik menggunakan Bahasa sarkasme seringnya menyeleweng dari aturan kesantunan Bahasa. Yang mengakibatkan kesalahpahaman diantara penutur dan mitra tutur karena dianggap kurang sopan santun [6]. Seta Kata-kata tidak baik dalam berbahasa Indonesia seringkali diucapkan tidak lain untuk menyerang kepada pihak tertentu, untuk mengungkapkan rasa kekecewaan, kekesalan atau bahkan meluapkan emosi[7]. tentunya jika dibiarkan akan menimbulkan kegaduhan, bahkan sampai perpecahan antar kelompok[8].

Bukan hal mudah untuk membendung hal ini, karena perlunya bimbingan serta arahan agar masyarakat khususnya netizen lebih teredukasi lagi untuk menjaga tutur bahasanya baik itu dalam, konten, informasi, berita, maupun sebuah penelitian.

Ucapan kasar juga ucapan menyakitkan perasaan yang berupa cacian dan sindiran sangat tidak dianjurkan oleh Rasulullah SAW. Karena dapat menimbulkan kebencian, permusuhan dan pertikaian. Bahkan dengan ucapan yang kotor mengatakan sindiran, sumpah serapah, kritikan tanpa ada dasar yang konkret merupakan ucapan yang harus di jauhi, karena melakukan hal itu seperti menganggap rendah lawan bicara. Begitupun dengan berkata tidak senonoh, hal ini sangat tidak menguntungkan untuk diri sendiri karena kata-kata yang tidak senonoh membawa energi negatif, yang jika dibiarkan dapat merusak kondisi mental dan pikiran kita[9]. Etika berbicara dijelaskan dalam Alquran :

قَوْلٌ مَّعْرُوفٌ وَمَغْفِرَةٌ خَيْرٌ مِّنْ صَدَقَةٍ يَتَّبِعُهَا أَدَىٰ ۖ وَاللَّهُ غَنِيٌّ حَلِيمٌ ﴿٥٦﴾

“Perkataan yang baik dan pemberian maaf lebih baik daripada sedekah yang diiringi tindakan yang menyakiti. Alah Mahakaya, Maha Penyantun.” (Al-Baqarah/2:265).

يَا أَيُّهَا الَّذِينَ آمَنُوا اتَّقُوا اللَّهَ وَقُولُوا قَوْلًا سَدِيدًا

“Wahai Orang-orang yang beriman! Bertakwalah kamu kepada Allah dan ucapkanlah perkataan yang benar.” (Al-Azhab/33:70).

Dalam hadits Al Imam Tirmidzi meriwayatkan, dimana Rasulullah SAW bersabda:” Sesungguhnya tidak ada sesuatu apapun yang paling berat ditimbangan kebaikan seorang mu, min pada hari kiamat seperti akhlak yang mulia, dan sungguh-sungguh(benar-benar) Allah benci dengan orang yang lisannya kotor dan kasar” (Hadits Riwaya At Tirmidzi nomor 2002, hadits ini hasan shahih, lafadz ini milik At Tirmidzi).

algoritma yang akan digunakan yaitu algoritma *Multinomial Naïve Bayes*, algoritma ini merupakan suatu metode pembelajaran *probabilistic* yang didasarkan teorema Bayes [10]. Keuntungan dalam algoritma *Multinomial Naïve Bayes* ini adalah memiliki performa yang baik serta tidak rumit[11]. Serta algoritma *Multinomial Naïve Bayes* ini cocok untuk diterapkan pada klasifikasi teks atau dokumen yang akan diolah dikarenakan algoritma ini sering disebut optimal dalam pengklasifikasian teks [12].

Penelitian yang berkaitan dengan text klasifikasi sudah banyak dilakukan oleh peneliti. Beberapa peneilitan sebelumnya yang berkaitan dengan Text klasifikasi seperti Tahun 2021 Yuyun dkk dalam jurnalnya mengenai algoritma *Multinomial Naïve Bayes* untuk klasifikasi sentimen pemerintah terhadap penanganan Covid-19 menggunakan data Twitter dengan 2000 dataset memiliki tingkat akurasi yang baik yaitu 74% [10]. Pada tahun 2020 Debby Alita dan Auliya Rahman membuat jurnal mengenai Pendeteksian Sarkasme pada proses analisis sentimen menggunakan *Random Forest Classifier* yang mengambil dataset dari Twitter berjumlah 2.027 dengan hasil rata-rata akurasi sebesar 16,61%[13]. Pada tahun 2020 Aripin dkk dalam jurnalnya mengenai identifikasi kelas emosi majemuk pada kalimat majemuk bahasa Indonesia menggunakan model Multinomial naïve bayes dengan tingkat akurasi yang cukup baik yaitu 75,47%[11]. Pada tahun 2020 Febry dan Addin dalam jurnalnya mengenai klasifikasi sentimen sara, hoaks dan radikal pada postingan media sosial menggunakan algoritma *Naïve Bayes Multinomial Text* yang mengambil data dari aplikasi Twitter berjumlah 260 dataset dengan Teknik pengambil data menggunakan Twitter API yang mehasilkan akurasi yang tinggi yaitu 94% [14].

Tahun 2019 Ahmad dkk dalam jurnalnya mengenai identifikasi konten kasar pada Tweet Bahasa Indonesia yang mengambil data dari Twitter juga dengan menggunakan algoritma klasifikasi *Multinomial Naïve Bayes* dan *Support Vector Machine* dengan *Linear Kernel*. Hasilnya algoritma *Support Vector Machine* dengan *Linear Kernel* lebih unggul yaitu dengan nilai akurasi 0.99 dan untuk algoritma *Multinomial Naïve Bayes* memiliki akurasi sebesar 0.98. Namun demikian keduanya memiliki performa yang sama baiknya[7]. Tahun 2019 Ivan dkk dalam jurnalnya mengenai klasifikasi *hate speech* berbahasa Indonesia di *Twitter* menggunakan *Naive Bayes* dan seleksi fitur *Informan Gain* dengan normalisasi kata. Data yang diperoleh sebesar 250 data dengan hasil akurasi terbaik yaitu 98%[2]. Tahun 2020 Umar dkk mengenai tentang implementasi *Multinomial Naïve Bayes* untuk klasifikasi ujaran kebencian pada dataset kicauan (*Twitter*) berbahasa Indonesia menggunakan melakukan konfigurasi untuk memodifikasi data *training* agar bisa mengatasi *imbalance* dataset dengan metode *oversampling* dan *random undersampling* yang menghasilkan akurasi paling bagus yaitu sebesar 0.53[12].

Perbedaan penelitian ini dengan penelitian sebelumnya adalah pada data dan topik yang diangkat, jumlah data yang digunakan berjumlah 5000 dataset yang 2500 berlabel sarkasme dan 2500 berlabel non sarkasme. Serta topik yang diangkat pada penelitian ini yaitu “sarkasme”. Serta penelitian ini diimplementasikan menjadi aplikasi pendeteksi sarkasme berbasis website.

Berdasarkan pemaparan tentang penelitian teks analitik bahwa data yang digunakan sebagian besar diambil dari twitter atau sekarang ini aplikasi X. Sarkasme juga dilakukan oleh penelitian sebelumnya dari Debby Alita dan Auliya Rahman mengenai Pendeteksian Sarkasme pada proses analisis sentimen tetapi menggunakan algoritma *Random Forest Classifier*. Sementara penelitian yang dilakukan menggunakan algoritma *Multinomial naïve Bayes*.

1.2. Rumusan Masalah

Berdasar latar belakang permasalahan tersebut, maka dapat dirumuskan permasalahan yang akan dikaji dalam penelitian ini yaitu sebagai berikut:

1. Bagaimana menerapkan *Multinomial Naïve Bayes* dalam sistem pendeteksi kata sarkasme pada aplikasi X?
2. Bagaimana mengukur tingkat akurasi dari *Multinomial Naïve Bayes* dalam sistem pendeteksi kata sarkasme pada aplikasi X?

1.3. Tujuan dan Manfaat penelitian

Tujuan yang ingin dicapai dalam melaksanakan penelitian ini adalah:

1. Implementasi sistem deteksi kata sarkasme menggunakan algoritma *Multinomial Naïve Bayes*.
2. Mengetahui tingkat akurasi algoritma *Multinomial Naïve Bayes* dalam mendeteksi kata sarkasme.

Adapun manfaat yang didapat dari penelitian ini adalah:

1. Mengedukasi penggunaan kata sarkasme dalam konten atau postingan
2. Melakukan pencegahan dan pendeteksian kata yang mengandung unsur sarkasme.

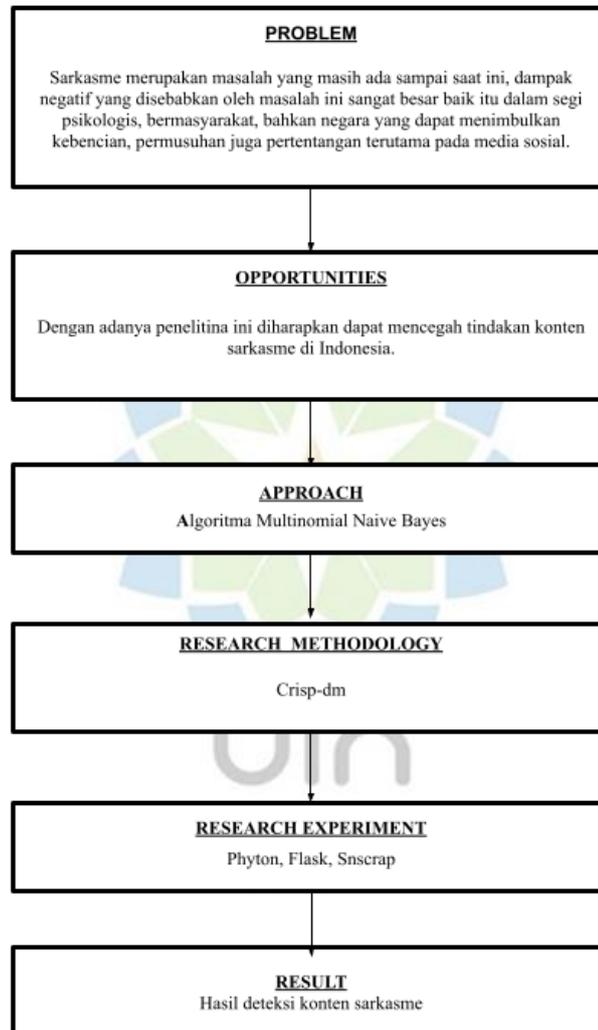
1.4. Batasan Masalah

Batasan masalah yang digunakan dalam suatu pembahasan diharapkan lebih terarah sesuai dengan tujuan yang akan dicapai. Adapun beberapa Batasan masalah tersebut meliputi:

1. Sistem deteksi menggunakan algoritma *Multinomial Naïve Bayes*.
2. Data yang digunakan dalam penelitian ini bersumber dari aplikasi X yang berbahasa Indonesia diambil menggunakan API dengan Teknik *Crawling*, serta dari beberapa website yang memuat contoh konten sarkasme.
3. Keyword: tolol, goblok, bangsat, bangke, bacot, pansos, ngemis-ngemis, anjing, babi, anjirr[15], muntah, jijik, mampus, becus, bodoh, dungu, sampah[16].
4. Jumlah data 5000, dengan rincian 2500 data sarkasme dan 2500 data non sarkasme

1.5. Kerangka Pemikiran

Kerangka pemikiran dari penelitian tugas akhir ini akan dijelaskan pada gambar berikut:



Gambar 1. 1 Kerangka Pemikiran

pada kerangka penelitian ini menggunakan menggunakan algoritma Multinomial Naive Bayes dikarenakan teknologi serta algoritma ini cocok untuk meneliti data berupa teks sertapenelitian ini menggunakan Metodologi penelitian Crips-dm. bahasa pemrograman menggunakan python serta tools yang digunakan yaitu flask dan untuk mengcrawling data menggunakan Snscrap.

1.6. Metodologi Penelitian

Pada penelitian ada beberapa tahapan yaitu pengumpulan data menggunakan API dengan Teknik *Crawling*, lalu tahapan Preprocessing untuk ekstrasi teks, setelah ekstrasi fitur dilanjutkan dengan klasifikasi, dan yang terakhir evaluasi.

1.7. Sistematika Penulisan

Pada penulisan laporan tugas akhir ini ada sistematika penulisan yang bertujuan agar sistematika penulisan dapat mudah dipahami, dan agar jelas arah serta tujuannya. Adapun sistematika penulisan dari penelitian ini yaitu:

BAB I PENDAHULUAN

Bab ini menguraikan latar belakang penelitian, rumusan masalah penelitian, Batasan masalah penelitian, tujuan penelitian dan manfaat penelitian, kerangka pemikiran, metodologi penelitian, dan sistematika penyusunan.

BAB II LANDASAN TEORI

Bab ini menjelaskan tentang teori-teori yang akan digunakan pada penelitian. Penjelasan tersebut meliputi Sarkasme hingga *Multinomial Naïve Bayes* sebagai algoritma yang digunakan pada penelitian ini.

BAB III METODOLOGI PENELITIAN

Bagian ini menjelaskan mengenai pendekatan yang diterapkan dalam penelitian ini, serta memberikan gambaran tentang langkah-langkah yang terlibat dalam setiap tahapan dari metode yang telah dipilih.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan isi dari laporan mengenai hasil penelitian, proses serta laporan penelitian, pembahasan akurasi hingga variable terkait penelitian.

BAB V PENUTUP

Bagian ini mencakup rangkuman dari keseluruhan laporan yang memberikan jawaban terhadap permasalahan dalam penelitian, serta memberikan rekomendasi dan usulan untuk penelitian mendatang.